CAS CS 599 B: Mathematical Methods for TCS

Lecturer: Mark Bun Spring 2022

Lecture Notes 10:

Bounded Independence Fools AC^0

Reading.

- Braverman, "Poly-logarithmic independence fools AC^0."
- See also: Tell, "The Bazzi-Razborov-Braverman Theorems" `https://sites.google.com/site/roeitell/Expositions`

Today, we'll explore the power of k-wise independent distributions to fool relatively complex functions. Recall:

Definition 1. A distribution $\mathcal{D}$ over $\{-1,1\}^n$ is said to ε-fool a function $f:\{-1,1\}^n \to \{-1,1\}$ if

$$|\mathop{\mathbb{E}}_{x\sim\mathcal{D}}[f(x)] - \mathop{\mathbb{E}}_{x\sim\mathcal{U}}[f(x)]| \leq \varepsilon.$$

We saw that k-wise independent distributions 0-fool degree-k polynomials, which include depth-k decision trees. How far can we push this? Depth-k decision trees are computed by size 2^k k-DNF and by k-CNF (i.e., DNF and CNF formulas where each bottom gate has arity k). Can we show that bounded independence fools small DNF/CNF? Or more ambitiously, small circuits of bounded depth?

Some intuition for why this may be true comes from the LMN polynomial approximation theorem that we saw when we studied the learnability of small circuits under the uniform distribution:

Theorem 2 (Tal's strengthening of LMN)**.** *Let f be computed by an $\{\wedge,\vee,\neg\}$ circuit of size s and depth d. Then there exists a polynomial $p:\{-1,1\}^n\to\mathbb{R}$ of degree $O(\log^{d-1}(s)\log(1/\varepsilon))$ such that $\mathbb{E}[(f-p)^2]\leq\varepsilon$.*

That is, every size-s depth-d circuit C is approximated by a $k = O(\log^{d-1}(s))$-degree polynomials, so we should expect any k-wise independent distribution to (approximately) fool C. Related observations led Linial and Nisan to conjecture, around 1990, that size-s depth-d circuits are fooled by $O(\log^{d-1}(s))$-wise independence.

It took about 20 years for this conjecture to turn into a theorem, which will be the main point of today's discussion:

Theorem 3 (Braverman's Theorem)**.** *Every size-s depth-d circuit is ε-fooled by every k-wise independent distribution with*

$$k = (\log s)^{O(d)} \cdot \log(1/\varepsilon).$$

Historical note: Essentially no progress was made on this question until 2007, when Bazzi [**?**] proved the result for DNF in a 53-page tour de force. Razborov then dramatically simplified the proof [**?**]. In 2009, Braverman proved the result for general AC^0 circuits obtaining the bound $k = (\log(s/\varepsilon))^{O(d^2)}$. This was subsequently improved by Tal [**?**] to $(\log(s/\varepsilon))^{3d+3}$ and by Harsha and Srinivasan [**?**] to the result stated above. Note that the latter two results are technically incomparable because of the constant hiding in the big-O. Linial and Nisan's conjectured bound of $O(\log^{d-1}(s)\log(1/\varepsilon))$ is unbeatable due to a counterexample by Mansour [**?**].

1 Sandwiching Polynomials

Given our intuition that low-degree approximability should correspond to approximate fooling by bounded independence, it's instructive to see why we can't immediately conclude Theorem 3 from Theorem 2. Suppose f is approximated in ℓ_2-distance by a degree-k polynomial, in that $\mathbb{E}_{x\sim\mathcal{U}_n}[(f(x)-p(x))^2] \le \varepsilon$ for p of degree k. We'd like to be able to show that $|\mathbb{E}[f(\mathcal{D})] - \mathbb{E}[f(\mathcal{U}_n)]| \le \mathrm{poly}(\varepsilon)$ for any k-wise independent $\mathcal{D}$. The natural thing to do is to try to break this up as

$$|\mathbb{E}[f(\mathcal{D})] - \mathbb{E}[f(\mathcal{U}_n)]| \le \underbrace{|\mathbb{E}[f(\mathcal{D})] - \mathbb{E}[p(\mathcal{D})]|}_{(1)} + \underbrace{|\mathbb{E}[p(\mathcal{D})] - \mathbb{E}[p(\mathcal{U}_n)]|}_{(2)} + \underbrace{|\mathbb{E}[p(\mathcal{U}_n)] - \mathbb{E}[f(\mathcal{U}_n)]|}_{(3)}.$$

Term (2) is zero by k-wise independence. Term (3) is at most $\sqrt{\varepsilon}$ by applying the triangle inequality and Cauchy-Schwarz:

$$|\mathbb{E}[p(\mathcal{U}_n)] - \mathbb{E}[f(\mathcal{U}_n)]| \le \mathbb{E}_{x\sim\mathcal{U}_n}[|f(x)-p(x)|] \le \mathbb{E}_{x\sim\mathcal{U}_n}[(f(x)-p(x))^2]^{1/2} \le \sqrt{\varepsilon}.$$

But we seem to be stuck with term (1). We want to be able to control this quantity for an arbitrary k-wise independent distribution, but one could be adversarially chosen to place most of its weight on the points where, say, $f(x) = 1$ and $p(x) = -1$.

One of Bazzi's key insights was to show that a stronger form of polynomial approximation rules out this possibility. Specifically, he showed that if f is approximated from above and below by a pair of "sandwiching polynomials" then it is fooled by bounded independence.

Lemma 4 (Bazzi)**.** *Suppose there are degree-k polynomials $p_\ell, p_u : \{-1,1\}^n \to \mathbb{R}$ such that*

1. $p_\ell(x) \le f(x) \le p_u(x)$ *for all* $x \in \{-1,1\}^n$*, and*
2. $\mathbb{E}_{x\sim\mathcal{U}_n}[p_u(x) - f(x)] \le \varepsilon$ *and* $\mathbb{E}_{x\sim\mathcal{U}_n}[f(x) - p_\ell(x)] \le \varepsilon$.

Then f is ε-fooled by k-wise independence.

Proof. As before, but without absolute values, we write

$$\mathbb{E}[f(\mathcal{D})] - \mathbb{E}[f(\mathcal{U}_n)] = \underbrace{\mathbb{E}[f(\mathcal{D})] - \mathbb{E}[p_u(\mathcal{D})]}_{(1)} + \underbrace{\mathbb{E}[p_u(\mathcal{D})] - \mathbb{E}[p(\mathcal{U}_n)]}_{(2)} + \underbrace{\mathbb{E}[p_u(\mathcal{U}_n)] - \mathbb{E}[f(\mathcal{U}_n)]}_{(3)}.$$

The second term is again 0, and the third term is at most ε. But now because p_u is an upper bound on f, we have that the first term is nonpositive. Hence

$$\mathbb{E}[f(\mathcal{D})] - \mathbb{E}[f(\mathcal{U}_n)] \le \varepsilon.$$

Similarly, we can use the lower sandwiching polynomial to show that $\mathbb{E}[f(\mathcal{U}_n)] - \mathbb{E}[f(\mathcal{D})] \le \varepsilon$. Thus, $|\mathbb{E}[f(\mathcal{D})] - \mathbb{E}[f(\mathcal{U}_n)]| \le \varepsilon$ for every k-wise independent $\mathcal{D}$. □

Bazzi's Lemma has a converse which gives a complete characterization of foolability by bounded independence. It can be proved using linear programming duality.

2 Approximating AC^0

Our goal now is to show that every small, low-depth circuit is approximated by a pair of sandwiching polynomials. It suffices to show that every such circuit has a "lower sandwich" since AC^0 is closed under negation: If p is a a lower sandwich for $-f$, then $-p$ is an upper sandwich for f.

Even though Bazzi's Lemma turns out to completely characterize foolability, there seems to be some wiggle room. For one, we really only needed the lower sandwich to satisfy $\mathbb{E}[p_\ell(\mathcal{D})] - \mathbb{E}[f(\mathcal{D})] < 0$; in particular, it may be easier to guarantee this weaker condition if we let p_ℓ depend on the distribution $\mathcal{D}$. Second, we don't necessarily need to sandwich f itself. By the triangle inequality, it would be enough to sandwich a *different* function f' which we know to be close to f under both the uniform distribution and under $\mathcal{D}$. These observations are captured in the following lemma.

WARNING: To break symmetry in discussing one-sided error approximations, we are switching from $\{-1, 1\}$ notation to $\{0, 1\}$ notation. The transformation $f(x_1, \dots, x_n) \mapsto g(y) = 1 - 2f(1 - 2y_1, \dots, 1 - 2y_n)$ shows that this changes nothing meaningful.

Lemma 5. *Let $f : \{0,1\}^n \to \{0,1\}$. Suppose that for every k-wise independent distribution $\mathcal{D}$, there exists a Boolean function $f' : \{0,1\}^n \to \{0,1\}$ and a degree-k polynomial p such that*

1. *f' approximates f under $\mathcal{D}$ and under the uniform distribution: $\Pr_{x\sim\mathcal{D}}[f(x) \neq f'(x)] \leq \varepsilon/3$ and $\Pr_{x\sim\mathcal{U}_n}[f(x) \neq f'(x)] \leq \varepsilon/3$.*
2. *p is a lower sandwiching approximation to f': $p(x) \leq f'(x)$ for all $x \in \{0,1\}^n$ and $\mathbb{E}_{x\sim\mathcal{U}_n}[f'(x) - p(x)] \leq \varepsilon/3$.*

Then $\mathbb{E}[f(\mathcal{U}_n)] - \mathbb{E}[f(\mathcal{D})] \leq \varepsilon$.

Proof.

$$\begin{aligned}
\mathbb{E}[f(\mathcal{D})] &\geq \mathbb{E}[f'(\mathcal{D})] - \varepsilon/3 \\
&\geq \mathbb{E}[p(\mathcal{D})] - \varepsilon/3 \\
&= \mathbb{E}[p(\mathcal{U}_n)] - \varepsilon/3 \\
&\geq \mathbb{E}[f'(\mathcal{U}_n)] - 2\varepsilon/3 \\
&\geq \mathbb{E}[f(\mathcal{U}_n)] - \varepsilon.
\end{aligned}$$

□

So to show that AC^0 is fooled by bounded independence, it is enough to show that every $f \in \mathrm{AC}^0$ is approximated by another function f' that has a lower sandwiching polynomial approximation.

Theorem 6. *Let f be computed by a size-s depth-d circuit. Then for every k-wise independent distribution $\mathcal{D}$, there exists a Boolean function $f' : \{0,1\}^n \to \{0,1\}$ and a degree-k polynomial p such that*

1. *f' approximates f under $\mathcal{D}$ and under the uniform distribution: $\Pr_{x\sim\mathcal{D}}[f(x) \neq f'(x)] \leq \varepsilon/3$ and $\Pr_{x\sim\mathcal{U}_n}[f(x) \neq f'(x)] \leq \varepsilon/3$, and*
2. *p is a lower sandwiching approximation to f': $p(x) \leq f'(x)$ for all $x \in \{0,1\}^n$ and $\mathbb{E}_{x\sim\mathcal{U}_n}[f'(x) - p(x)] \leq \varepsilon/3$,*

where $k = (\log s)^{O(d)} \cdot \log(1/\varepsilon)$.

The proof of Theorem 6 combines two different kinds of polynomial approximations. The first is the ℓ_2 LMN approximation of Theorem 2. The second is a so-called "probabilistic polynomial" approximation.

Lemma 7. *For every size-s, depth-d circuit f and every distribution $\mathcal{D}$ over $\{0,1\}^n$, there exists a polynomial p with $\Pr_{x\sim\mathcal{D}}[p(x) \neq f(x)] \leq \varepsilon$ such that*

1. $\deg p \leq (\log s)^{O(d)} \cdot \log(1/\varepsilon)$
2. $\max_{x\in\{0,1\}^n}\{|p(x)|\} \leq \exp((\log s)^{O(d)} \cdot \log(1/\varepsilon))$
3. *There exists a size-*$\mathrm{poly}(s)$*, depth-$O(d)$ circuit E such that $p(x) \neq f(x) \iff E(x) = 1$.*

Constructions of these kinds of polynomials go back to work of Razborov, Smolensky, and Tarui in the late 80's and early 90's. They were, for example, used to show that the majority function is not in AC^0. The key observation in Braverman's work was that these polynomials satisfy condition 3: That the "error region" on which such a polynomial disagrees with the circuit it's approximating can itself by detected using a small circuit. A construction achieving the stated parameters of Lemma 7 was given by Harsha and Srinivasan **??**.

Proof. Apply Lemma 7 using distribution $\frac{1}{2}(\mathcal{D} + \mathcal{U}_n)$ and error parameter $\varepsilon/8$. This gives a polynomial p_0 such that $\Pr_{x\sim\mathcal{D}}[f(x) \neq p_0(x)] \leq \varepsilon/4$ and $\Pr_{x\sim\mathcal{U}_n}[f(x) \neq p_0(x)] \leq \varepsilon/4$, as well as a size-$\mathrm{poly}(s)$, depth-$O(d)$ circuit E such that $f(x) \neq p_0(x) \iff E(x) = 1$.

Now invoking Theorem 2, let p_E be the ℓ_2 approximation to circuit E of degree $\log(s)^{O(d)} \cdot \log(1/\delta)$ such that $\mathbb{E}[(E - p_E)^2] \leq \delta$.

Finally, set $f' = f \vee E$, $q = p_0(1 - p_E)$, and our final approximating polynomial $p = 1 - (1 - q)^2$. Intuitively, we are taking p_0 to be our "base" approximation to f', with the understanding that p_0 occasionally makes wild errors when $E(x) = 1$. Multiplication by the polynomial $(1-p_E)$ serves to mollify these errors, but may not result in a sandwiching approximation when $f'(x) = 1$. Setting $p = 1 - (1 - q)^2$ forces the final polynomial to be upper bounded by 1.

Let us now check formally that the conditions of Theorem 6 are satisfied for an appropriate choice of δ.

1. Under either the uniform distribution or under $\mathcal{D}$, we have $\Pr[f(x) \neq f'(x)] \leq \Pr[E(x)] = \Pr[f(x) \neq p_0(x)] \leq \varepsilon/4$.

2. To show that p is a lower sandwiching approximation to f', we'll first show that q is a "one-sided error" approximation to f. That is:

   Claim 8. $f'(x) = 0 \implies q(x) = 0$.

   To see this, suppose $f'(x) = 0$. Then $E(x) = 0$, so we are not in the mistake set, and hence $p_0(x) = f(x) = 0$. So $q(x) = 0$ as well.

   Claim 9. $\|f' - q\|_2 \leq \sqrt{\varepsilon/4} + \exp(\log(s)^{O(d)} \cdot \log(1/\varepsilon)) \cdot \sqrt{\delta} \leq \sqrt{\varepsilon/3}$ *by taking* $\delta = \varepsilon \cdot \exp(-\log(s)^{O(d)} \cdot \log(1/\varepsilon))$.

   To see this, we use the triangle inequality to bound

   $$\|f' - q\|_2 \leq \|f' - p_0(1 - E)\|_2 + \|p_0(1 - E) - q\|_2.$$

   For the first term,

   $$\|f' - p_0(1 - E)\|_2 \leq \sqrt{\Pr[E(x) = 1]} \leq \sqrt{\varepsilon/3}.$$

For the second term, we write $p_0(1-E) - q = p_0(p_E - E)$. Using the fact that $\max_x\{|p(x)|\} \leq \exp(\log(s)^{O(d)} \cdot \log(1/\varepsilon))$, we get

$$\|p_0(1-E) - q\|_2 \leq \exp(\log(s)^{O(d)} \cdot \log(1/\varepsilon)) \cdot \|p_E - E\|_2 \leq \exp(\log(s)^{O(d)} \cdot \log(1/\varepsilon)) \cdot \sqrt{\delta}.$$

We'll now use these two claims to show that p is a lower sandwiching approximation. First, it is a pointwise lower bound on f' by the following reasoning. If $f'(x) = 0$, then by Claim 8, we have $q(x) = 0$ so $p(x) = 0$. On the other hand, if $f'(x) = 1$, then $p(x) \leq 1$ by construction.

Now to show that p approximates f', we observe that:

- If $f'(x) = 0$, then $p(x) = 0$ so there is no error.
- If $f'(x) = 1$, then $f'(x) - p(x) = (1 - q(x))^2 = (f'(x) - q(x))^2$.

Hence, $\|f' - p\|_1 \leq \|f' - q\|_2^2 \leq \varepsilon/3$.

Finally, note that the degree of p is $(\log s)^{O(d)} \cdot \log(1/\varepsilon) + (\log s)^{O(d)} \cdot \log(1/\delta)$ where $\delta = \exp(-(\log s)^{O(d)} \cdot \log(1/\varepsilon))$. This gives a total degree bound of $(\log s)^{O(d)} \cdot \log(1/\varepsilon)$. □